

# A Semantic Rule-based Framework for Efficient Retrieval of Educational Materials

Maryam Tayefeh Mahmoudi<sup>1, 2</sup>, Fattaneh Taghiyareh<sup>1</sup>, Kambiz Badie<sup>2</sup>

<sup>1</sup>School of Electrical and Computer Eng., College of Eng., University of Tehran, Tehran, Iran

<sup>2</sup>Knowledge Management & E-Organizations Group, IT Research Faculty, Cyber Space Research Institute (Ex. Research Institute for ICT), Tehran, Iran

[mahmodi@itrc.ac.ir](mailto:mahmodi@itrc.ac.ir)

[ftaghiyar@ut.ac.ir](mailto:ftaghiyar@ut.ac.ir)

[k\\_badie@itrc.ac.ir](mailto:k_badie@itrc.ac.ir)

**Abstract:** Retrieving resources in an appropriate manner has a promising role in increasing the performance of educational support systems. A variety of works have been done to organize materials for educational purposes using tagging techniques. Despite the effectiveness of these techniques within certain domains, organizing resources in a way being adequately reusable for support purposes is still in the offing. In this paper a semantic approach is proposed to increase performance of retrieving educational materials based on using frames. Here, frames are used to represent the very knowledge necessary for realizing the similarity/ relevance between query and supportive materials. Owing to the complexity in semantic handling of the entire text, the suggested frame-based approach is applied only to the titles or subtitles, or in general the main headings, in the material. To make these frames comprehensive, we have made use of two attributes called "Major Characteristics" and "Basic Constituents", which are responsible respectively for "the goal behind a conceptual entity (Why/ for What a conceptual entity is being used)" and "the basic elements supporting a conceptual entity (How / in What way a conceptual entity is realized)". Conceptual entities here stand for the entire ideas belonging to headings (titles or subtitles) in a material. These attributes seem to have enough potential for representing the knowledge of titles and sub-titles in a way reflecting the content of the paragraphs in a reasonable way. To evaluate the capability of the proposed approach, retrieving materials within the domain of Multi-Agent Systems (a subject of high concern in Artificial Intelligence) was picked out as the benchmark problem. According to this benchmark, materials are retrieved based on the user's desire. Taking this point into account, we made a dataset for the subject of Multi-Agent Systems as an educational resource in academia, within which a number of users' desires from different groups were considered as possible queries, and the corresponding materials were then retrieved using the proposed approach. Computer experiments show acceptable precision and recall values for these queries with a quite good balance between them which is represented in terms of F-measure. . The findings lead us to the fact that "Major Characteristics" and "Basic Constituents" have the ability to increase the status of re-usability for the stored materials. Moreover, the fact that materials can be reused efficiently, leads us to the point that our proposed representation scheme can be useful for educational support issues in the situations where user's desire is complex enough to the extent that several materials ought to be merged together to yield the requested material.

**Keywords:** Semantic retrieval, material retrieval/ reuse, educational materials, frame-based representation, frame attribute, major characteristics, basic constituents

## 1. Introduction

Nowadays, increasing digital resources, including educational materials, have made a great challenge for innovating techniques to solve the problem of finding the truly-necessary information. It is interesting to notice that, when it comes to educational support systems, facilitating the way that learners get their appropriate resources becomes remarkably important.

Within the above scope, many annotating, tagging and indexing techniques have been created for this purpose. These techniques generally make use of WordNet and meta-data as appropriate knowledge representation schemes to describe the resources (Zhao et al, 2008) (Kohler et al 2006) (Dobsa, 2007) (Roy et al, 2008). Also, Latent semantic indexing (LSI) and concept indexing (CI) are among those techniques which are capable of organizing educational assets and offer effective search and categorization services (Dobsa, 2007)(Gómez et al, 2004). While these methods improve the detection of relevant documents on the basis of the terms found in queries, there also exist

some query reformulation techniques whose effort is to make mapping between tagging and querying vocabularies (Bischoff et al, 2010) in an acceptable manner.

In sum, statistical & semantical techniques are the two major categories that however belong to indexing, annotating and tagging (Moreda et al, 2007) (Zhang et al, 2011). It should be noted that, despite the advantages & effectiveness of these techniques within certain domains, they suffer from their own limits and deficiencies particularly when it comes to organizing resources with regard to inadequacy of reusability for support purposes.

To overcome these deficiencies, in this paper, we introduce a frame-based semantic technique to enhance the retrieval mechanism for educational resources. The proposed technique is, not only capable of resolving the existing ambiguity in tags, but can also embed the focal knowledge for exploring the similarity relevance between query and supportive materials. The frames used in the suggested technique are benefited by two attributes called "Major Characteristics" and "Basic Constituents" which stand respectively for "the goal behind a concept" and "the elements that support a concept to be realized" (Mahmoudi et al, 2004) (Badie et al, 2008). These attributes seem to be comprehensive enough to reveal the knowledge behind headings of a material as well as the status of learner's query. With regard to this, improving the semantic ability of retrieving educational resources through Major Characteristics (MJ) and Basic Constituent (BC) becomes the prime concern of our paper.

The rest of the paper is organized as follows: Section 2 reviews some of the previous works that have been done in the areas of indexing, annotating and organizing educational resources. Section 3 describes our proposed approach, while, in Section 4, experimental results are analyzed. Section 5 includes conclusion and future works.

## **2. Related works**

Mankind is facing too many information resources such as: tutorials, books, learning materials, reports, case studies and practices, etc. which are to be used for educational purposes. Either organizing and retrieving appropriate educational resources or classifying them are thus major issues in learning environments. Within this context, there exist various systems based on semantic retrieval that are capable of organizing educational assets. Content management (Shao et al, 2003), information retrieval (Liu et al, 2008), question-answering (Moreda et al, 2011), classification (Thorlechter et al, 2013), recommender (Zheng et al, 2011), educational support and intelligent tutoring (Günel et al, 2010) systems, can also be enumerated as means for this purpose. These systems are equipped with various types of semantic approaches such as annotating, tagging and indexing, that may facilitate resource organization process.

One approach to such issue is to make use of ontology-type structures such as WordNet and metadata which is in fact a way to describe the resources in a neat and efficient way (Lohmann et al, 2008), while tag and time are mostly useful in predicated user's preference and recommending related resources. It is to be noticed that a tag performs as a bridge between a user and a resource through which user's preference for the desired resource is expressed, and the more frequently a tag is used, it means that the more the user is interested in the related resource (Zheng et al, 2011). Generating tags in a collaborative way that is called *folksonomy* (Bateman et al, 2007) from the one side and using tags and tags clouds to discern credible content in online message forums (Grady et al 2012) from the other side, can also have a significant role in annotating and categorizing resources, especially for adaptable online learning purposes. Apart from simple annotation methods, there also exist some co-constructed semantic space for information fusion which exploits effective annotation (Lee et al, 2012). Besides, some light-weight techniques and tools such as Cerno have been proposed

for legacy code analysis and mark-up towards semi-automatic semantic annotation of textual documents according to a domain-specific semantic model (Kiyavitskaya et al, 2009).

Indexing has also a significant role in retrieving and processing the educational contents and resources (Mahmoudi et al, 2011). Various algorithms, approaches and networks are applied for indexing purposes such as: Latent Semantic Indexing (LSI) (Thorleuchter et al, 2013), Enhanced Instance Retrieval Network (Lourenço et al 2010), and terminologies (Dinh et al 2012). In all these cases, the ultimate goal is to specify accurately the resource's content, in order to elicit words which are semantically related to this content.

Statistical models, natural language processing, multi-label classifiers, and collaborative techniques are in the meantime most commonly used for tag recommendation (Alepidou et al, 2011). It is obvious that bridging the gap between tagging and querying vocabularies can also yield improving the potential of resource organization systems (Bischoff et al, 2010).

Web query analysis for different application domains using semantic and linguistic knowledge have also the ability to illustrate how far a higher number of relevant resources can be retrieved (Conesa et al, 2008). Syntax-based query reformulation (SQR) and query cluster summarization (QCS) have in the meantime the ability to enhance the performance of information retrieval (Lioma et al, 2008) in this regard. There is no doubt that structured document retrieval (SDR) leads to a better retrieval performance in terms of both precision and functionality especially for textual resources (Liu et al, 2008). Semantic roles extracted from natural language texts have also been shown to be important for improving the semantic information performance of question answering systems (Moreda et al, 2011). Moreover, corpus-based approaches, that make use of statistical models to determine the semantic role of constituents of a sentence, have also been shown to be useful for both information retrieval and question answering purposes (Moreda et al, 2007).

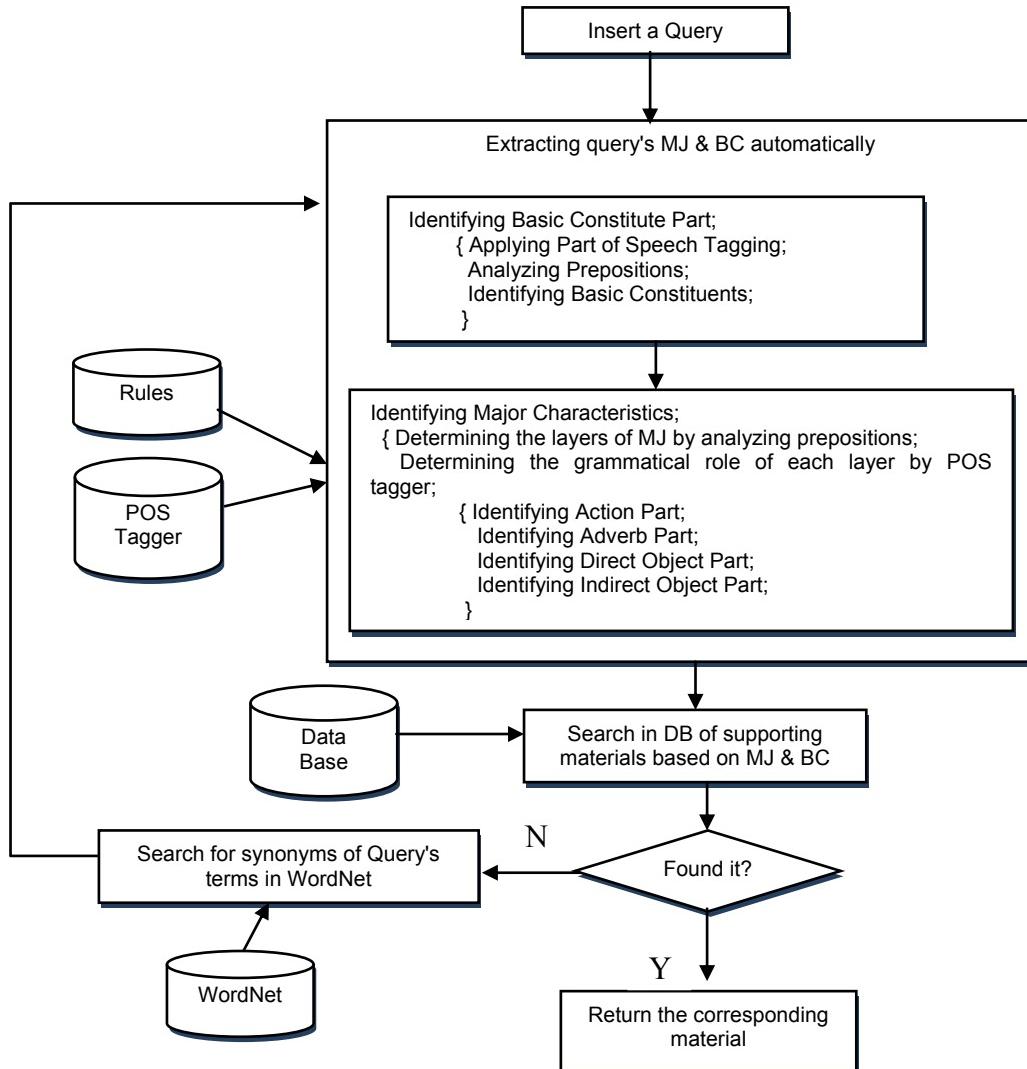
As a conclusion, to increase the speed and efficiency of education supportive systems as well as to have a flexible and reusable repository of e-learning materials, it would be crucial to perform annotation of the document with special metadata in a way as automatic as possible (Roy et al, 2008). Although the mentioned approaches to annotating, tagging and indexing are widely used in different search, retrieval and text processing applications, they still suffer from deficiencies in their semantic potentials. To enhance their semantic capabilities, structures such as frames with comprehensive attributes and values may help a lot. In this respect, determining informative attributes as those we have proposed, seems to be an appropriate solution and capable enough to reveal the main purposes behind phrases. . In this regard, enhancing the semantic ability of the retrieval process and reusability of education supportive materials through using frames with particular attributes are our main concern in this paper.

### **3. The proposed approach**

#### **3.1 Basic idea**

As it was mentioned before, in large-scale databases of education supportive materials, using a well-defined semantic approach plays a significant role in retrieving desired resources. In this respect, we propose a frame-based semantic retrieval approach that seems to be capable enough to facilitate such a process. Figure 1 illustrates the details of the proposed framework.

As illustrated in Figure 1, each time a query is presented; it should first be grammatically analyzed and be compared with the "titles" of the existing supporting materials within the data base. For this purpose, extracting the title of supporting material and its parsing are necessary, in order to figure out the values of corresponding "Major Characteristics" and "Basic Constituents" as the major attributes.



**Figure 1:** Details of the proposed framework

Figure 2 illustrates the pseudo code of our suggested approach. Here, semantic rules as well as frames can be used to realize such a process. Frames are helpful in the sense of determining grammatical roles of the existing terms in queries and titles with regard to education supportive materials in a database. Two significant attributes that are considered for this purpose are "Major Characteristics (MJ)" and "Basic Constituents (BC)". "Major Characteristics" is the attribute which explains the main objective behind using a material, while "Basic Constituents" mainly focuses on the methods, techniques or tools which are used to realize this objective (Badie et al, 2008)(Badie et al, 2004).

```
Inserting a query;  
Parsing the query;  
Extract (MJ, BC) of query;  
{ Identifying Basic Constitute Part;  
{ Applying Part of Speech Tagging;  
Analyzing Prepositions;  
Identifying Basic Constituents; }  
Identifying Major Characteristics;  
{ Determining the layers of MJ by analyzing prepositions;  
Determining the grammatical role of each layer by POS tagger;  
{ Identifying Action Part;  
Identifying Adverb Part;  
Identifying Direct Object Part;  
Identifying Indirect Object Part;  
}  
}  
}  
Search in Database of supporting materials' titles (MJ, BC);  
If (query (MJ, BC) = supporting materials' titles (MJ, BC));  
Send existing response based on relatedness of query (MJ, BC) to title (MJ, BC);  
Else  
Finding synonyms of terms of query in WordNet;  
Repeat the whole process to extract appropriate supporting materials for that;
```

**Figure 2:** The pseudo code of suggested approach

Having studied several titles within supporting materials, we acquired some rules, which were subsequently used to extract "Major Characteristics" and "Basic Constituents" from a "title". Certain conjunctions and prepositions can be in charge of specifying the values of these attributes. (Mahmoudi et al, 2011). It has been found out that some propositions being used in a conjunction, like "in" and "for" followed by a verb usually yield two layers for "Major Characteristics", which follow the same structure including four main parts of "Action", "Adverb/Adjective", "Direct object" and "Indirect object". "Action part" is mainly a verb, while, "Direct object" is a noun or a pronoun that becomes subject to this verb or shows the result of the related action. It is able to answer "What?"s or "Whom?"s relating to this verb. In addition to the grammatical roles, "Indirect object" is also the recipient of the "direct object" and has the ability to answer "To whom?"s or "For whom?"s and it usually follows a preposition. The last part belongs to "Adverb/ Adjective", which can modify verbs, adjectives, clauses, sentences, and other adverbs. It typically answers "How?"s, "In what way?"s, "When?"s, "Where?"s, and "To what extent"s, etc.

For BCs, most of the time, one layer at maximum seems to be sufficient. Determining BC is therefore closely related to the conjunctions which are considered for this purpose. Some of these conjunctions are "based on", "on the basis of", "on the ground of", "using", "making use of", "taking into", etc (Mahmoudi et al, 2011).

It is to be noticed that, employing a POS tagger can facilitate the process of determining grammatical role of terms in query as well as supporting materials. For example, consider the title "Considering agent mobility architecture for controlling transportation based on FIPA standards". As it was mentioned, the terms coming after "based on" would stand for BC, while those coming before "based on" stand for MJ. It is to be noted that Major Characteristics in this example includes two

layers which are separated by "for". The grammatical role of each part and their components are represented in Figure 3.

**Major Characteristics (MJ)**

	Action-part	Adverb-part	Direct Obj.	Indirect Obj.
Layer 1 (MJ <sub>1</sub> )	Considering	-	agent mobility architecture	-
Layer 2 (MJ <sub>2</sub> )	controlling	-	transportation	-

**Basic Constituents (BC)**

1 <sup>st</sup> Layer	2 <sup>nd</sup> Layer
FIPA standards	-

**Figure 3:** Major characteristics & basic constituents for the example of "considering agent mobility architecture for controlling transportation based on FIPA standards"

Having reviewed large amount of titles, several semantic rules are yielded for distinguishing MJs and BCs, For example:

- IF a "word" or "phrase" comes after "via" or "based on" THEN it is most probably a BC.
- IF the rest of the title consisting of a verb comes after "for" THEN it is most probably a MJ's 2nd layer.

The same rules can be applied both for analyzing the queries and the titles as well. After extracting MJs and BCs in a query, based on the rules discussed above, the process of searching for the learning materials (whose "MJ"s and "BC"s are identical to those in the query) would become subject to performance. Having found the corresponding terms, the related material will then be retrieved. Otherwise closely related words or synonyms from WordNet have to be extracted for the same purpose. For the moment a WordNet with simple relational structures has enough potential to respond successfully to our study. Retrieving process will continue in this way, and in the cases where no related term was found, the process will be terminated and a failure notice will then be issued. Types of conjunction/preposition and their status with regard to the attributes, and Rules distinguishing MJs and BCs are illustrated respectively in Table1 and Table2.

**Table1:** Types of conjunction/proposition and their statuses with regard to the attributes

Name of Attribute	Possible Types of Conjunctions or Propositions	Status with regard to the highlighted attribute
Major Characteristics	With the purpose of	With the purpose of X
	With the aim of	With the aim of X
	In order to	In order to X
	With the objective of	With the objective of X
	For/ in	Y for/in X (sentence with Action part)
Basic Constituents	Based on	Y Based on X
	On the basis of	On the basis of X
	Using/ via/ by	Y Using/via/by X
	On the ground of	On the ground of X
	Making use of	Making use of X

**Table2:** Some rules distinguishing "MJ"s and "BC"s

IF	THEN	X for Y(Y is non-action type noun)	Y is Adverb part in the value of MJ
		X for/in Y (Y is a phrase including action part)	Y is second layer value of MJ
		X-based/-oriented/-inspired Y	X is Adverb & Y is Direct Object in the value of MJ
		X of Y(Y is non-action type noun)	Y is Indirect Object in the value of MJ
		X based on/via/by/using Y	Y is Basic constituents

As is seen from Table1, where for instance we say "with the purpose of X", "with the aim of X", "in order to X", ... in a heading, we are in some way showing that "X" is the value of "Major Characteristics" of the entire heading. Also, when we have "Y based on X", "on the basis of X", "Y using/via X", ... in a heading, we intend in some way to demonstrate that "X" is worth being propounded as the value of "Basic Constituents" in the entire heading. Let us keep in mind that the basic motive for extracting such values is to see what the main purpose of the entire heading is, and what constituents it is grounded on. Coming to Table2, when for instance we say "X for Y", if "Y" is a noun then it is regarded as something describing "X" and thus a part of the value of "Major Characteristics", while if the same "Y" is a phrase including action part, it would be regarded as a second layer value of "Major Characteristics". Also, for instance if we have "X of Y" in the heading, provided that "Y" is a non-action type noun, "Y" would then be regarded as an indirect object.

#### 4. Assessment of the proposed Approach

##### 4.1 Experimental set-up

In order to evaluate the proposed approach, we made a data set including 134 supporting materials in the domain of Agent Science and Technology. To perform our tests, we designed some questions to help finding proper materials as the answers for these questions. Results were evaluated through comparing the responses made by our approach with the real responses obtained from experts in the domain of Agent Science & Technology.

##### 4.2 Experimental Requirements

To apply frames to headers, we first split the header into the corresponding "Basic Constituents" and "Major Characteristics". Based on the information obtained in such a manner it is determined whether the value of "Major Characteristics" holds one layer or two layers. Here, it is essential to find out the grammatical role of the terms included in the values of "Major Characteristics" as well as those included in the value of the "Basic Constituents".

To realize the grammatical roles of the terms, we may make use of rules to decide what role a term can hold. For instance, to realize an "action-part" (in the value of "Major Characteristics"), rules can take into account the information regarding suffixes like "tion", "sion", "ment", "ing", etc. which are linguistically significant. Let say, if, for example, POS tagger is tagging a word without "tion" as a verb, and that word to be not located between two nouns, one may conclude that the term must be an "action". An example for applying such a rule is illustrated in Figure4.

To identify the "adverb" part, we make use of adverb and adjective tags already produced by Stanford POS tagger.

##### 4.3 Analysis of the Results

To evaluate our approach, a dataset comprising of supportive materials was used. .Results demonstrate the fact that our proposed approach has the potential to function well with regard to detecting the values of "MJ"s and "BC"s, included in the titles. It should be mentioned that out of the 134 titles used in our experiments, our approach has been able to function properly in 107 cases. As we shall show, the amount of precision in detecting the values of "MJ"s and "BC"s is 93%, while the same amount for recall is 79%.

"Simulation of Dialogue Management"
Simulation
Simulate → is a verb
Not between two nouns → is an action
"Reliable transactions in multi-agent systems"
Transaction
Transact → is a verb
Not between two nouns → is an action

**Figure 4:** Samples of a rule-based approach for identifying the action part

In addition to the above experimentation, we designed some questions that cover a large variety of possible cases. Respecting this, both "verbs" and "objects" were given as the input, and the names of "appropriate materials" were then returned as the output. Table3 illustrates the detailed information regarding each question.

As the final stage in evaluation, we made use of "F-measure" to totally assess the efficiency of our approach. "F-measure" is indeed for the purpose of measuring the very essential balance which is to exist between "recall" and "precision". To determine "F-measure" we need both "precision" and "recall" values. "Precision" is measured as the proportion of relevant retrieved documents to the number of retrieved documents, while "Recall" is measured as the proportion of relevant retrieved documents to the total number of relevant documents. In the meantime F-measure shows the harmonic mean of these two functions.

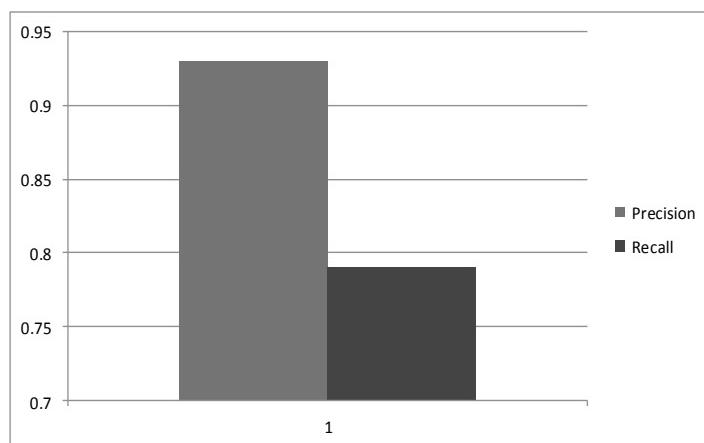
Experimental results reveal that the value of precision belonging to our approach is equal to  $(31/33) = 0.93$ . Also, the value of recall was found to be  $31 / (31+8) = 0.79$ , since eight materials were left out (Figure5). Taking these two values into account, the value of "F-measure" was determined in the following way.

$$F\text{-measure} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

$$= 2 * (0.93 * 0.79) / (0.93 + 0.79) = 0.85$$

**Table3:** Precision & recall of some queries

Query			Automatic approach					Non-automatic Approach	
			Precision	Recall	Relevant Retrieved	Retrieved	Relevant	Precision	Recall
Description	Relation-determination	What is the relation between communication and agents?	1	1	4	4	4	1	1
		What is the relation between protocols and agents?	1	0.5	2	2	4	1	1
		What is the relation between web and agents?	1	1	6	6	6	1	1
	Causality-determination	What are the reasons of deception in networks of mobile sensing agents?	1	1	1	1	1	1	1
	Outcome-determination	What is the outcome of Engineering of multi-Agent?	1	0.33	1	1	3	1	0.66
	Historical-determination	What is the history of multi-agent systems?	1	1	2	2	2	0.5	0.5
		How was the evolution of applications with agents?	1	0.5	1	1	2	1	1
	Definition	What is the definition of multi-agent systems?	0.5	0.4	2	4	5	1	0.8
		What is the definition of an assistance agent?	1	1	2	2	2	1	1
Inclusion	Comparison	What are the differences between soccer-playing intelligent robots with a multi-agent system and regular one?	1	1	1	1	1	1	1
	Solution-Determination	What are the multi-agent solutions in power engineering applications?	1	0.3	1	1	3	1	1
	Role	What is the role of ontology?	1	1	1	1	1	1	1
	Application	What is the application of multi-agent in knowledge retrieval?	1	1	5	5	5	1	0.8
		What is the usage of conceptual maps?	1	1	1	1	1	1	1
	Advantages& disadvantages	What is the usage of learning automata?	1	1	2	2	2	1	1



**Figure 5:** Precision & recall of the proposed approach

As it is understood from the achieved F-measure, our proposed approach has been capable of avoiding irrelevant results. That is because of following a specific pattern in MJ and BC parts for each type of question. In fact, we tried to design some patterns that avoid producing irrelevant results. One should however not forget the fact that applying WordNet itself has also been helpful in this regard.

## 5. Concluding remarks

In the paper, we demonstrated how attributes called "Major Characteristics" and "Basic Constituents" can be used to realize the process of semantic retrieval of education supportive materials in an efficient way. These attributes were shown to be potential enough for representing the knowledge belonging to titles and sub-titles in a way reflecting paragraphs content in a reasonable way. Here, some rules based on a variety of grammatically significant propositions and conjunctions, were used to detect the values of "Basic Constituents" and "Major Constituents" in the titles and subtitles. Rules can be constituted based on the existing linguistic knowledge. It however should be noted that the higher number of propositions in a rule, a higher expectation may exist with regard to its effective role in detection. For the moment to avoid extra computation, rules have been decided to include only a few predicates. However, developing more potential rules through considering further predicates and applying complicated thesaurus to match alternative words can be regarded as major research works for future. This calls for further analysis of the existing titles as well as sub-titles in the existing materials with the purpose of discovering a wide range of conjunctions and prepositions as essential requirements, for constituting adequate rules.

As the final point, it should be noted that the suggested approach to retrieval of supportive materials can be adopted as a popular approach to retrieval due to its ability in processing texts (textual information) with no particular emphasis on using natural language processing, which can be both complicated and time-consuming in nature.

## References

- Alepidou,Z.I., Vavlakakis, K. N. and Mitkas, P.A. (2011) "A Semantic Tag Recommendation Framework for Collaborative Tagging Systems", IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, pp 633-636.
- Badic, K. and Mahmoudi, M.T (2008) A Computational Framework for Manipulating an Issue from the View-Point of Other Issues, 14th Intl. Cong. of Cybernetics and Systems of WOSC – ICCS'08 ,Wroclaw, Poland, pp 9-12.
- Badic, K. and Mahmoudi, M.T. (2004) An approach to generating new ideas based on linking frames of concepts, ECCBR'04, Computational creativity workshop, 7th European Conf. In Case-based reasoning, Madrid, Spain.

- Bateman,S., Brooks,C., McCalla, G. and Brusilovsky.P.(2007) Applying Collaborative Tagging to E-Learning, In the Proceedings of the Workshop on Tagging and Metadata for Social Information Organization, Banff, Canada.
- Bischoff,K., Firan, C.S.,Nejdl,W. and Paiu, R.(2010) " Bridging the gap between tagging and querying vocabularies: Analyses and applications for enhancing multimedia IR", Web Semantics: Science, Services and Agents on the World Wide Web, Vol 8, pp 97-109.
- Conesa,J., Storey,V.C. and Sugumaran, V. (2008) "Improving web-query processing through semantic knowledge", Data & Knowledge Engineering, Vol 66, No.1, pp 18-34.
- Dinha, D., Taminea, L., Boubekeurb, F. (2012) "Factors affecting the effectiveness of biomedical document indexing and retrieval based on terminologies", Artificial Intelligence in Medicine, Article in Press.
- Dobsa, J. (2007) "Comparison of information retrieval techniques: Latent semantic indexing (LSI) and Concept indexing (CI)", Solomonovi seminarji, Faculty of Organization and Informatics, Varazdin, University of Zagreb.
- Gómez, J. M., Cortizo, J.C., Puertas, E. and Ruiz, M. (2004) "Concept Indexing for Automated Text Categorization", Natural Language Processing and Information Systems, Lecture Notes in Computer Science, Vol 136, pp 495-502.
- Grady, L. O., Wathen, C.,N., Burger, J., C., Betel, L., Shachak, A., Luke, R.,Hockema, S., Jadad, A., R. (2012) " The use of tags and tag clouds to discern credible content in online health message forums", International Journal of Medical Informatics, Vol 81, pp 36-44.
- Günel, K. and Aşlıyan, R.(2010) "Extracting Learning Concepts from Educational Texts in Intelligent Tutoring Systems Automatically", Expert Systems with Applications, Vol 37, issue 7, pp 5017-5022.
- Kiyavitskaya, N., Zeni, N, Cordy, J., R., Mich, L., Mylopoulos, J. (2009) "Cerno: Light-weight tool support for semantic annotation of textual documents", Data & Knowledge Engineering, Vol 86, pp 1470-1492.
- Kohler,J.,Philippi, S., Specht, M. and Ruegg, A.(2006) "Ontology based text indexing and querying for the semanticweb" , Knowledge-Based Systems, Vol 19, pp 744–754.
- Lee,C.H, Wang, S.H.(2012) "An information fusion approach to integrate image annotation and text mining methods for geographic knowledge discovery", Expert Systems with Applications, Vol 39, pp 8954-8967.
- Lioma,C. and Ounis, I.(2008) "A syntactically-based query reformulation technique for information retrieval", Information Processing and Management, Vol 44, No.1, pp 143-162.
- Liu, S., McMahon, C.A. and Culley, S.J. (2008) "A review of structured document retrieval (SDR) technology to improve information access performance in engineering document management", Computers in Industry, Vol 59, pp 3–16.
- Lohmann ,S., Thalmann ,S. , Harrer A. and Maier,R.(2008) Learner-Generated Annotation of Learning Resources – Lessons from Experiments on Tagging, Proceedings of I-KNOW '08 and I-MEDIA '08, Graz, Austria, pp 304-312.
- Lourenço, A., Carreira, R., Peña, D.G., Méndez, J.R., Carneiro, S., Rocha, L.,M., Díaz, F., Ferreira, E., C., Rocha, I., Riverola, F., Rocha, M. (2012) "BioDR: Semantic indexing networks for biomedical document retrieval",Expert Systems with Applications, Vol 37,pp 3444–3453.
- Mahmoudi, M.T., Taghiyareh, F., Rajavi, K. and Pirouzi, M.S. (2011) A Context-Aware Framework for Semantic Indexing of Research Papers, The Fourth Int'l. Conf. on Information, Process, and Knowledge Management (eKnow 2011), Valencia, Spain.
- Mahmoudi, M.T. and Badie, K. (2004) Content determination for composite concepts via combining attributes' values of individual frames, IKE'04, Intl. Conf. On Information & Knowledge Engineering, Las Vegas, USA.
- Moreda, P., Llorens,H., Saquete,E.and Palomar,M. (2011) "Combining semantic information in question answering systems" ,Information Processing and Management, Vol 47, pp 870–885.
- Moreda, p., Navarro,B. and Palomar, M. (2007) "Corpus-based semantic role approach in information retrieval" Data & Knowledge Engineering Vol 61, pp 467-483.
- Roy,D., Sarkar,S. and Ghose, S.(2008) "Automatic Extraction of Pedagogic Metadata from Learning Content", I. J. Artificial Intelligence in Education, Vol 18, No.2, pp 97-118).
- Shao, N.W.Y., Yang, S.J.H. and Sue, A.Y.S. (2003) A content management system for adaptive learning environment. Multimedia Software Engineering. Proceedings. Fifth International Symposium on Multimedia Software Engineering (ISMSE'03), pp 206-214.
- Thorlechter, D., Poel, D.,V. (2013) " Technology classification with latent semantic indexing", Expert Systems with Applications, Vol 40,pp 1786–1795.
- Zhang, W., Yoshida, T. and Tang, X. (2011) A comparative study of TF\*IDF, LSI and multi-words for text classification, Expert Systems with Applications, Vol 38, No.3, pp 2758-2765.
- Zhao, N., Fang,F. and Fan,L.(2008) "An Ontology-based Model for Tags Mapping and Management", International Conference on Computer Science and Software Engineering, pp 483-486.
- Zheng, N. and Li,Q.(2011) "A recommender system based on tag and time information for social tagging systems", Expert Systems with Applications, Vol 38,No.4, pp 4575-4587.